Exploring and characterizing the dispersion of wages in Algeria, using the two methods "principal components analysis and hierarchical ascending clustering "

ADJOUT Abdelkader *

Labo-Financial and banking systems and macroeconomic policies facing international fluctuations / Hassiba Ben Bouali University of Chlef (Algeria), a.adjout@univ-chlef.dz

Received :25/09/2023 **Accepted :13/01/2024** **Published :19/01/2024**

Abstract: *This study mainly aims to compare the temporal evolution from 2009 to 2019 of the monthly average net wages received by workers in their various categories of qualifications, within the various economic activities in the legal, public and private sectors in Algeria.*
This comparison, which the study aims at, is based on a multidimensional statistical analysis of the dispersion of wages in Algeria, using two integrated methods represented in the analysis by the principal components and the hierarchical ascending clustering according to the Ward algorithm. After the two processes of analyzing the data, using the statistical software R, and interpreting the results of the two methods, it was concluded that 18 dimensions representing the variables of the study were summarized into two dimensions that represent two principal components, which we were able to name as the factor of average activity wages for the study period and the factor of the extent of wage dispersion between executives and enforcement agents within the activity. According to these two factors, the clustering process produced five different groups of activities.
Key words*: Wage dispersion, principal component, hierarchical clustering.*
Jel Classification Codes : C38, J24, J3

Introduction

The issue of wages in all its dimensions has taken an important place among theorists of schools of economic thought on the one hand, and for official and unofficial, international, regional and national organizations and bodies on the other hand, that's because it's of great importance related to the daily lives of individuals and societies, and the economic policies adopted by states and governments, as well as decision-makers within companies and institutions. One of the dimensions that has been and continues to be of interest to them is the subject of the determinants of wage dispersion, in which various theories have been presented.
To apply models of the theories presented and to test their validity or the validity of parts of them in a country's economy, most of them require detailed data mainly related to national wage surveys, whether in institutions and companies, or in individuals and households. These surveys are carried out mainly in Algeria by the National Office of Statistics every year starting in 2009, the last of which was in 2020 among a sample of national public and private institutions and companies, and one survey among Algerian households in 2011 on consumer spending and the standard of living of households in Algeria, part of the questionnaire of this survey was devoted to wage (ONS, 2014, p. 2). Unfortunately, the National Office of Statistics provides analysis of the results of these surveys only, and does not provide raw data for survey questionnaires after collection. Perhaps this is the main reason why there are no studies on the application of one of the models of theories of wage dispersion in Algeria comprehensively, except for a few studies, which relied on the results of these surveys or relied in their studies on wage data for one or two institutions to apply one of the econometric models according to the collected data, and try to

* Corresponding author

identify, analyze and explain the wage gap, whether between sector activities, between institutions, between gender, between educational levels or ... Etc.

The National Office of Statistics (ONS) reports on wage differentials in Algeria for each survey it conducted with a sample of national public and private institutions and companies from 2009 to 2019. The content of the reports submitted is important to understand the extent of wage dispersion between economic activities, whether in the public or private sector, or the size of wage dispersion between permanent and non-permanent workers, or between qualification's categories... Etc. Since these reports are mainly addressed to the government and decision-makers to help them make the right decisions on the one hand, and to all components of society that are interested in this subject, even if they are not specialists, we find that ONS relies in its analysis on one-dimensional descriptive statistics tools, like other reports of this kind submitted by various bodies inside Algeria or from other countries.

Based on this, we decided to present an exploratory and descriptive analysis of wage dispersion in Algeria for the period 2009-2019, using two complementary methods, namely the Principal Component Analysis (PCA) and Hierarchical Ascendant Clustering (HAC) method, in order to answer the following problematic: **How are wages distributed in time between the economic activities of both the national public and private sectors and according to qualification's categories? Is it possible to form activity groups that differ clearly from each other along the study period?**

Study hypotheses: To answer the problematic, we rely on the following two hypotheses:

- Wages are distributed chronologically between activities at an increasing pace, but in a way that is close to monotony in terms of rates of increase. - More than two clearly different groups of activities can be formed throughout the study period.

Objectives and methodology: The overall aim of the study is to compare the monthly net average wages of different economic activities within the legal, public and private sectors. In order to achieve this, and to answer the problematic of the study, we used the descriptive analytical methodology.

Previous studies: As mentioned above, there are very few previous studies that have been presented to analyze the determinants of wage dispersion in Algeria, but in spite of that, we mention:

- A scientific article by Mohammed Boutouba and Ahmed Touil (2022) entitled: "Wage's gap between economic sectors in Algeria" (Boutouba Mohamed T. A., 2022, pp. 273, 297). The researchers used cross-sectional time series models to apply them to the wages of economic activities for the public and private sectors from 2000 to 2019 for three categories of workers (Executives, control agents, and enforcement agents), and concluded the following results:
  - There is a long-term wage gap for the public sector, while the private sector did not show accurate results.
  - The wage gap between activities is due to the number of labor force and the profitability of the worker, except for activities of a service nature, such as: hotels, restaurants, and social services.

- Master's thesis by Siraj Wahiba (2008) entitled: "An econometric study of wage fairness in Algeria" (Wahiba, 2008). The researcher used a regression model containing dummy variables to estimate the differences between the wages received by workers with their three qualifications (Executives, control agents, and enforcement agents) within Sonelgaz and a cement factory for the year 2007 separately, i.e. vertical comparison, and between the two companies, i.e. horizontal comparison. Finally, the researcher reached the following results:
  - The wages of workers with the same level as the workers of the cement company are higher than those paid to Sonelgaz workers.
  - There are differences between the wages of workers within the same organization but they are reasonable.

- Increasing productivity in the cement company leads to an increase in real wages, but the rate of increase in wages was less than the rate of increase in productivity.

➢ A scientific article by Mouloud Nourine and Mohamed Mahmoud Mohamed Aissa (2020) entitled: "Factors affecting the relationship between education and wages in the service sector, an econometric study applied to Algeria Telecom in Mostaganem" (Nourine Mouloud, 2020, pp. 135, 154). The researchers relied on multiple linear regression to estimate the "Mincer" earning model, based on the data of a questionnaire distributed to a sample of employees at Algeria Telecom in Mostaganem. After estimating the model and conducting all the necessary tests, they had reached the following conclusions:

- Increasing the number of years of schooling by one year leads to an increase in the monthly wage by 5.2%.
- The effect of experience is positive on the monthly wage, but this effect decreases in the long run.
- The return at the secondary level was very weak, estimated at 0.83%.
- The return at the university level was weak and decreasing, reaching -0.31%.

1. Theoretical framework for wage dispersion: On the subject of wage dispersion, various theories have been presented, starting with the theory of compensating wage differentials, as this theory states that perfect competition in various labor markets will inevitably lead to heterogeneity in wages among workers, which in turn results only from differences related to the difficulty or hardship of the tasks assigned to the worker and his skills (Pierre Cahuc, 2001, pp. 171-178).

When it comes to the dispersion associated with the hardship of tasks, the explanatory theory is the "Hedonic wage theory", first established by Adam Smith at the end of the eighteenth century, and recently it has become a full-fledged theory thanks to S. Rosen (1974). The summary of this theory is that the employee is assigned to accomplish a set of tasks of different difficulty and hardship, which are reflected in the pace of work to be accomplished, the work environment, the risks of accidents, and the social status that the worker benefits from. Rosen's proposal takes into account the heterogeneity of wages resulting from these "compensatory differentials", as this theory shows that perfect competition mechanisms lead to compensation for the most hardship work, and thus receive higher wages.

Skill pay differentials are explained by Gary Stanly Becker's Human Capital Theory (1964), who argues that investments in learning can be profitable, because learning results in an accumulation of skills, which amount to human capital, so the return from it should lead to higher wages. Overall, this theory aims to explain wage differentials through differences in individual productivity and investments in training and learning, which individuals make throughout their lives. On the other hand, the acquisition of valuable skills is accompanied by the costs of learning and training in the labor market. So it's like investments in physical assets that will yield a return in the future. These costs include: expenses for studies, possible losses of income resulting from the fact that the time spent on study was not devoted to paid activity, and psychological costs related to possible stress and inconvenience associated with learning and training. Becker distinguishes between public human capital, which improves individual productivity for all jobs, and private human capital, which only raises productivity for an exceptional type of job. In practice, public human capital is characterized mainly by the level of certification and professional experience in the labor market (Mincer, 1958, 1974), while private human capital reflects the career path of the employee within each company, and can be measured by seniority in the organization (INSEE, 2006, p. 30).

Both theories, despite their importance, assume the conditions of perfect competition in the labor market, these hypotheses are far from reality, so labor market theorists were not satisfied with them, but tried to explain the dispersion of wages by presenting other theories under the conditions of incomplete competition, such as monopolistic competition, for example. Perhaps

the most prominent of these theories are the theory of labour market segmentation, and the theory of efficiency wages. The first is the general framework in which the binary analysis of wages falls, which was founded in the seventies by "Doeringer" and "Piore". According to this approach, the labor market is not considered a homogeneous market, but instead is analyzed as adjacent parts, distinct from each other in the specific ways in which each part in order to access it, to work and to practice activity in it. The basic model presented proceeds from the fact that the labor market is a dual market, consisting of two parts, a primary sector that provides prestigious jobs, and a secondary sector that offers other precarious, poorly protected, and lower-paid jobs. It is also characterized by weakness of trade unions, unfavorable working conditions compared to the primary market, and no prospects for a viable career (Tchibozo, 1998, p. 24).

The second prominent theory in its various forms, which does not support the assumptions of perfect competition in the labor market, is theory of efficiency wages. Different approaches to this theory have been formulated by many neo-Keynesians, beginning with Leibentien (1957), who put forward this concept in his approach, which came against the traditional assumption of the relationship between wage and marginal productivity of the worker. According to his perspective and those who came after him within the framework of this current in explaining the dispersion of wages and everything related to the labor market in general, the marginal productivity of the worker depends mainly on his wage and not vice versa.

Leibentien's conclusion is that offering wages allows a person to live better and thus become more productive. He also concluded that the concept of effort is a key element in the relationship between productivity and pay. Among the models presented to analyze and explain this relationship within the framework of this theory are the shirking model of Shapiro et Stiglitz (1984), the model of reducing the costs of turnover of employee by Salop (1979), the equity model of Akerlof (1982), and the adverse selection model of Weiss et Landau (1984) (Teiva, 2010, p. 10).

2. Submission of data

We took data from the results of the national surveys of wages in institutions and companies conducted by the National Office of Statistics every May from 2009 to 2019. The background information gathered through the questionnaires relates to employees according to the qualification categories of permanent and non-permanent wage earners, who worked in the survey month (May), where the remuneration included the amounts of the basic salary, bonuses and allowances, gross salary and net salary received (ONS, 2010, p. 1).

The data that make up the preliminary data table, whose coding and designation are detailed in Appendix 1 are:

- 19 statistical individuals, each representing an economic activity (branch) within the national legal sectors, public and private, according to the Algerian nomenclature of activities currently in force, with the exception of agriculture and administration.
- The active variables are the monthly net average wages in Algerian dinars for each qualification category (executives, control agents, and enforcement agents) for the years 2009, 2011, 2013, 2015, 2017, 2019.
- To enrich and enhance the analysis, we included 12 quantitative and qualitative variables as supplementary variables, which do not contribute to the formation of the factor axes, but are placed by projection after the completion of finding these axes that make up the factorial planes. supplementary quantitative variables are the ratios of wages for a particular activity to the total average wages of a qualification category at the beginning and end of the study period (2009 and 2019), and the ratios of wages for a particular qualification category to the total average wages of an activity at the beginning and end of the study period (2009 and 2019). The supplementary qualitative variable is the legal sector, which takes two modalities, public or private.

3. Description of the problem

The overall objective of the study is to compare the monthly net average wages of different economic activities within the legal, public and private sectors.

3. 1/ From the perspective of lines (activities within the legal sectors): Each activity within the public or private sector is described by 18 average monthly net wages. From the perspective of this angle, what activities are similar to each other? Which of them differ? In general, Can we make a typology of activities, highlighting the similarities and dissimilarities between them? Assuming this typology is made, can we determine whether the similarities (or dissimilarities) relate to sectoral proximities (or divergences) (public or private) on the one hand, and (or) to proximities (or divergences) either in the temporal evolution of the qualification's categories shares from the average wages of these activities? or (and) the temporal evolution of activities' shares from the average wage of one or more qualification classes? So the goal of studying activities is to analyze their variability.

3. 2/ From the perspective of the columns (variables: average wages of qualification's categories in different years): Each category of qualification in a given year is seen through the monthly averages of the net wages of 19 individuals representing the various activities according to the Algerian Nomenclature of Activities on the one hand, and according to the legal sector on the other. The nature of the problem is to compare the distribution of these activities (from highest paid to lowest-paid) between two or three different categories of qualification (Executives, control agents, and enforcement agents) in one or more years of the study period. These comparisons between qualification's categories are made in different years according to the concept of linear correlation between them, where if two different categories are in the same year or in two different years that are more related, this means observing approximately the same distribution of 19 activities according to their average wages, and the same observation in the case of whether these two categories are for the same qualification in two different years. The opposite is true, they are weakly correlated, if the highest (or lowest) wages for both are not found in the same activities.

All of the above leads to the following questions: What are the qualification categories for the same year or a different year associated with each other? Which of them is weakly correlated to each other? In general, can we make an assessment of the linear correlations between 18 variables and read it directly from their graph, represented by correlation circle? Are average monthly net wages correlated to the legal sector? On the other hand, if the qualification's categories in different years are correlated, is there redundancy in the information given by the 18 columns of the initial data table in some form? Can this information be summarized by substituting 18 variables with a few artificial variables?

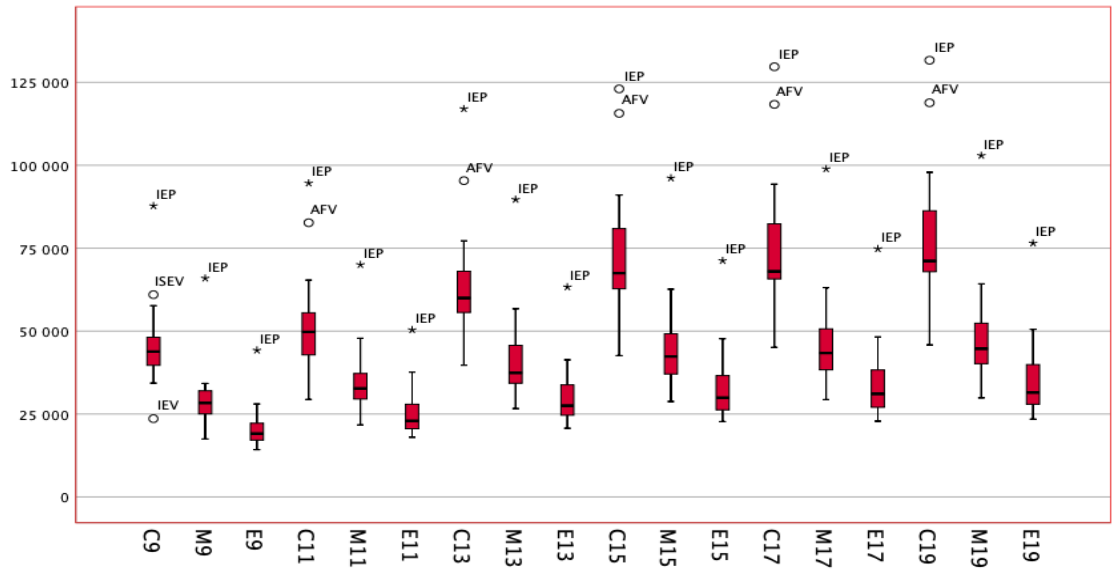
3. 3/ The role of supplementary variables: We discussed the problem that the wages of the qualification's categories should be analyzed taking into account the position of the legal sector, so we added this variable that has two modalities, public and private. We also added 12 quantitative variables, which we explained earlier, in order to determine their position on the factor planes that will be taken to interpret the results, and then help us to characterize the axes ، and if possible, name the factors (artificial variables).

4. Data Inspection

This step is necessary before starting to apply any statistical method, which researchers often overlook (Joseph F. Hair JR, 2019, p. 45). This is mainly to look for missing values, identify outliers, test the hypotheses on which most multivariate techniques are based, and address all of these problems, if any, in appropriate ways. The data we analyze doesn't contain any missing value, so we'll just look for and treat outliers. These values, defined for example by Barnett et Lewis (1994), as a value (or set of values) that appear to be inconsistent with the rest of the data (Planchon, 2005, p. 20). To detect outliers, we will use the boxplot of the active variables, in order to observe the values that exceed the boundaries of the arms of the box (the moustache), i.e: that exceed $Q_3 + 1{,}5(Q_3 - Q_1)$from above, and exceed $Q_1 - 1{,}5(Q_3 - Q_1)$from below,

where: Q_1: the first quartile (lower quartile), Q_3: Third quartile (upper quartile) , $(Q_3 - Q_1)$: interquartile range.

Figure 1: Graphical representation of active variables before winsorizing

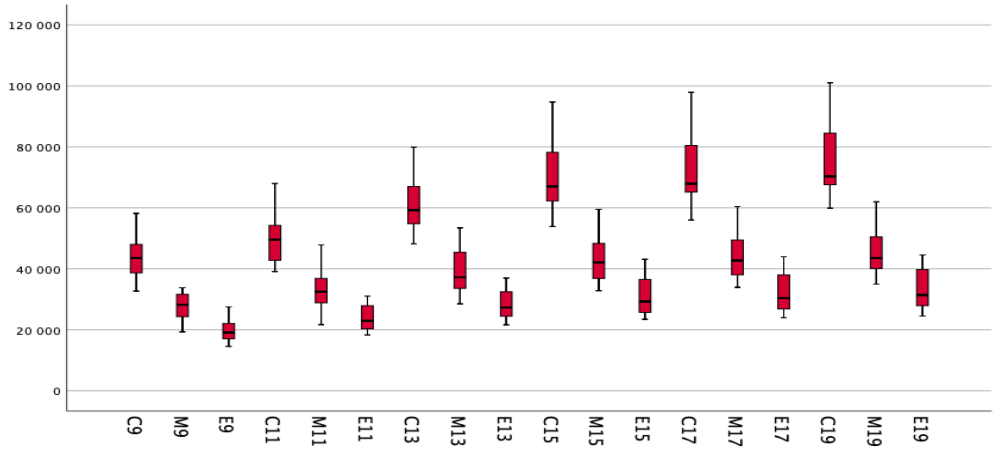


Source: In-person setup based on the initial data table and using SPSS software

All points outside the boundaries of the arms of the boxes representing the variables are outliers (Figure 1), where we see that the activity "IEP" in all variables represents an outlier, with the activity "AFV", but to a lesser degree, and once for the activities "IEV" and "ISEV" for the variable "C9". To treat these outliers, we used Winsorizing on the R program. However, some outliers remained, especially those related to IEP activity (Appendix 2), so we have seen that the best solution to address this problem is to take the IEP as a supplementary individual, we only care about its position on the factorial planes, without contributing to the formation of the factor axes, and we winsorize the data another time using the R program. After this process, no outliers remained, as shown in Figure 2, as there is no point outside the boundaries of the arms of all the boxes.

We note here that we applied the principal component analysis method before winsorizing and after the first winsorizing, so the individual "IEP" actually appeared a very extreme value, as we found before winsorizing that this individual contributes alone 66.36% to the formation of the first factor axis, which explains 86.12% of the total variance of the point cloud, and after the first winsorizing we found that it contributes alone 38.81% to the formation of the first factor axis, which explains 77.41% of the total variance of the point cloud.

Figure 2: Graphical representation of active variables after the final winsorizing



Source: In-person setup based on the initial data table and using SPSS software

5. Analysis and discussion of results

5. Analysis and interpretation of the results of the application of PCA to the data

5. 1/ Sample suitability test: To measure the suitability of the sample taken for analysis, the Kaiser-Meyer-Olkin (KMO) scale is used with a value between 0 and 1, where a weak value indicates that composite factors cannot be extracted, or in other words, does not allow the sample to produce the pattern we previously imagined.

Table 1: Measurement of sample suitability

Variable	E13	M13	C13	E11	M11	C11	E9	M9	C9	KMO Overall Value
KMO Value	0,61	0,62	10,5	90,	0,65	50,6	20,8	0,69	70,4	
Variable	E19	M19	C19	E17	M17	C17	E15	M15	C15	
KMO Value	0,54	0,6	0,41	0,58	0,7	80,3	0,66	0,56	0,39	**0,6**

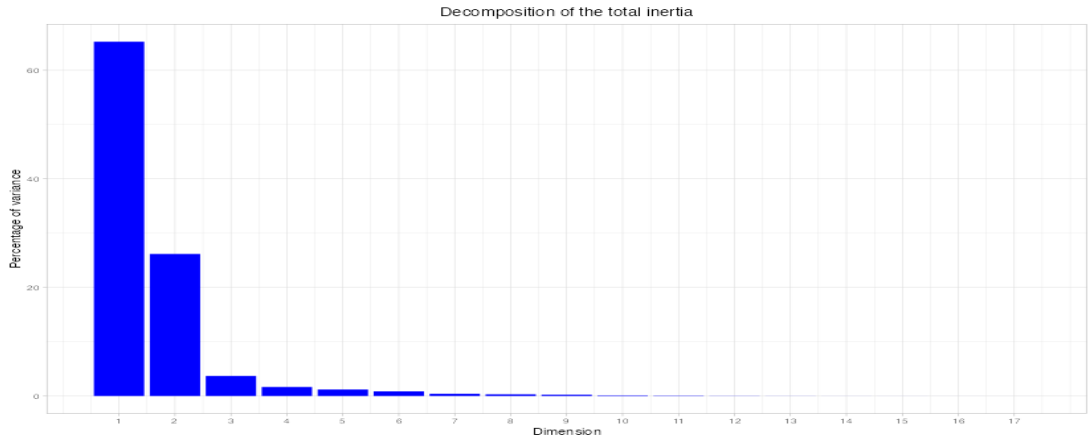
Source: Output of applying the PCA method to the data table, using the XL-STAT software

The values of this scale for variables (Table 1) range separately from 0.39 to 0.9, and most are greater than 0.6, and the overall KMO value is 0.6, which is mediocre but acceptable. Thus, the study sample is suitable for factor analysis.

5. 2/ Indicators of total variance (inertia):

5. 2.1/ Eigenvalues and the proportions of variance explained by each axis: In the standard principal component analysis method, the total variance of the individual point cloud (or variables) is equal to the number of active variables.

Figure 3: Scree plot



Source: Output of applying the PCA method to the data table using the R program

The variance explained by the first factor plane is 91.42% of the total variance (Appendix 3), where the first axis alone largely dominates the total ratio, with an eigenvalue of 11.75, corresponding to 65.26% of the total variance. It is followed by the second axis with an eigenvalue of 4.71, corresponding to 26.16% of the total variation. The rest of the factorial axes explain only small percentages of total variance, knowing that all eigenvalues related to them are less than 1, and according to the "Kaiser" criterion we do not take them into account in interpreting the results, and regardless of this criterion or other criteria, we will only interpret the results on the first factor plane (F1,F2), because 91.42% is very sufficient to explore the data table by representing the activities and qualification's categories for different years on it.

Table 2: Criteria to assist in the interpretation of activities on the factorial plane (F1, F2)

Activities	coordinates		contribution		(Cos2)		
	F1	**F2**	**F1**	**F2**	**F1**	**F2**	**(F1,F2)**
IMP	2,92	-1,42	4,03	2,39	0,73	0,17	13
EGEP	-0,63	-1,44	0,19	2,44	0,12	0,65	0,77
CP	-0,51	-0,46	0,12	0,25	0,12	0,1	0,22
CRP	3,36	-2,08	5,35	5,1	0,67	0,25	0,92
HRP	-2,42	-1,04	2,78	1,27	0,59	0,11	0,7
TCP	6,62	-0,94	20,74	1,04	0,96	0,02	0,98
AFP	2,97	-1,63	4,17	3,14	0,71	0,21	0,92
ISEP	-0,34	-3,18	0,05	11,9	0,01	0,82	0,83
SCSP	0,28	-1,46	0,04	2,5	0,02	0,52	0,54
IEV	-5,84	-1,19	16,15	1,68	0,94	0,04	0,98
IMV	-2,03	0,93	1,95	1,01	0,7	0,15	0,85
CV	-4,83	-0,44	11,03	0,23	0,97	0,01	0,98
CRV	-1,31	3,14	0,81	11,66	0,12	0,72	0,84
HRV	-1,54	0,32	1,13	0,12	0,52	0,02	0,54
TCV	-2,84	-0,63	3,83	0,47	0,82	0,04	0,86
AFV	7,58	2,62	27,16	8,09	0,87	0,1	0,97
ISEV	-0,6	4,47	0,17	23,61	0,02	0,87	0,89
SV	-0,83	4,43	0,33	23,1	0,03	0,88	0,91
IEP (supplementary)	27,17	2,14	/	/	0,98	0,01	0,99

Source: Output of applying the PCA method to the data table using the R program

5. 2.2 / Contribution of individuals to the formation of factorial axes: The formation of factor 1 (F1) is mainly due to four activities: Financial activities in the private sector "AFV", transport and communications in the public sector "TCP", extractive industries in the private sector "IEV", construction in the private sector "CV", which explain 75.08% of its variation.

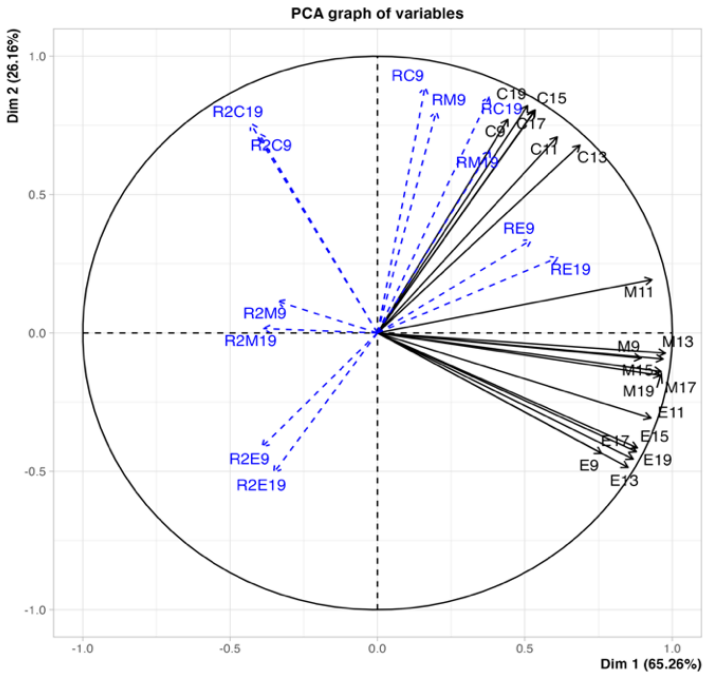
The second factor (F2) is mainly due to the activities: real estate, rental and services to enterprises in the private sector (ISEV), health in the private sector (SV), real estate, rental and services to enterprises in the public sector (ISEP), trade and repair in the private sector (CRV), which explains 69.56% of its variance (its Inertia).

5. 3/ Interpretation of factor 1

5.3.1/Coordinates of active variables:

The coordinate of a variable on the axis of a given factor in the standard principal component analysis method represents the linear correlation coefficient between this factor and this variable. From the table of coordinates of variables (Appendix 4), we can see that all variables are positively linearly correlated with the first factor, and in terms of intensity, most of them are strong except for the executives category (C9,...,C19) for all years, its correlations are medium in general (from 0.44 to 0.69), and all these variables are positively correlated with each other, where we can see from the correlation circle (Figure 4) that all active variables are on the right side of F1, and all measurements of angles that can be formed by any two vectors of two different variables are confined between zero and $\pi/2$. This means that the first factor represents a "size factor", and the explanation for the size effect in general is that if a variable takes a large value for a particular individual, all other variables also take large values, and vice versa (Ludovic Lebart, 2006, p. 94). In other words, the size effect expresses that some individuals have large values for all variables, and others have small values for all variables (Brigitte Escofier, 2008, p. 35).

Figure 4: Graph of variables (correlation circle)



Source: Output of applying the PCA method to the data table using the R program

In our study, this indicates that some activities, both in the public and private sectors, are characterized by higher monthly wages for workers than others throughout the study period and for all categories of qualification. In other words, if the wage of a qualification's category in a given year for a particular activity is large, all the wages of the qualification categories in that same year and all the wages of the three qualification's categories in the other years of that activity also take large monthly wages.

5.3.2 / Coordinates of activities (individuals): By virtue of the fact that the first axis is a size factor, and based on the transition relations that link the coordinate of a given individual on a given factor axis of individuals "$F_\alpha(i)$" with all the coordinates of variables on the factor axis of variables with the same rank, or vice versa, linking the coordinates of a given variable on a given factor axis of variables "$\phi_\alpha(j)$" with all the coordinates of individuals on the factor axis of individuals with the same rank. Mathematically, the two relations are written as follows:

$$\begin{cases} F_\alpha(i) = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{j=1}^{P} \frac{x_{ij} - \bar{X}_j}{\sigma_j} \phi_\alpha(j) \dots\dots\dots 1 \\ \phi_\alpha(j) = \frac{1}{n} \frac{1}{\sqrt{\lambda_\alpha}} \sum_{i=1}^{n} \frac{x_{ij} - \bar{X}_j}{\sigma_j} F_\alpha(i) \dots\dots\dots 2 \end{cases}$$

where: $\boldsymbol{F_\alpha(i)}$: the coordinate of individual i (i=1,...,n) on the factor axis F of the order α

$\boldsymbol{\phi_\alpha(j)}$: the coordinate of the variable j (j=1,...,P) on the factor axis **φ** of the order α

$\boldsymbol{x_{ij}}$: the value of the variable **j** taken by the individual **i**. $\bar{X}_j$: mean of variable **j**

$\boldsymbol{\sigma_j}$: Standard deviation of variable **j**. **n**: number of individuals. **P**: number of active variables.

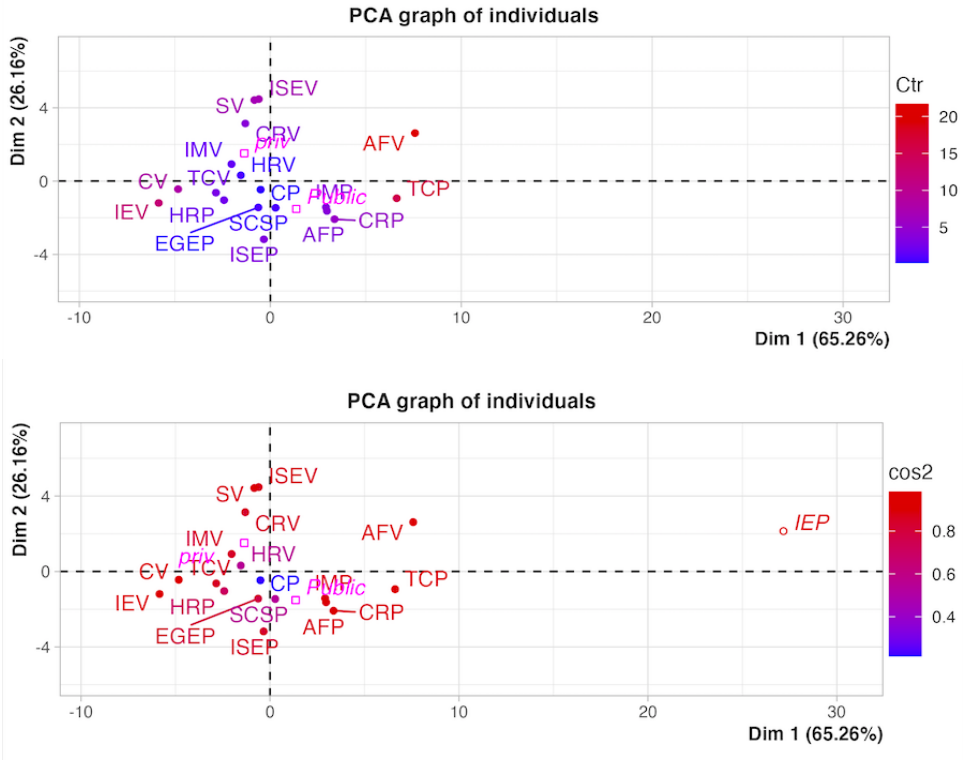
λ_α: the "λ" eigenvalue related to the factor axis of "α" order.

We will inevitably find that the activities with high monthly wages are positioned to the right of the first factor axis, i.e. take positive coordinates, the highest paid take the largest positive coordinates, and activities with low monthly wages are positioned to the left of this axis, i.e. take negative coordinates, and the least paid take the lowest coordinates on it. Indeed, this is what we observe on the graph of the variables (correlation circle). The first factor axis is an opposition axis of the highest-paying activities of all qualification's categories, in particular financial activities in the private sector "AFV" , transport and communications in the public sector "TCP" (in addition to the activity of extractive industries in the public sector "IEP", which is located at the far end of this axis. Although it is an supplementary variable that does not contribute to the formation of factor axes, it is very characterized by this axis, which is the highest wage activity for all qualifications throughout the study period), and the lowest paid activity for all qualification's categories, in particular extractive industries in the private sector (IEV) and construction in private sector (CV), which are located far left.

Thus, whatever the qualification's category in any year of the study period, the average net monthly wage in any of the AFV and TCP activities is greater than the total average wage for all activities, and the average net monthly wage in either IEV and CV activity is lower than the total average wage for all activities. It should be noted here that the lowest paid activity does not mean that it takes the lowest wage value compared to other activities throughout the period and whatever the qualification. The reverse is true for the highest paid activity (we exclude the IEP because we put it as an supplementary variable, so whenever we refer to the highest paid activity, we mean the activity "AFV", because it is an active variable, but it remains the second highest paid activity), and this is because the first transition relations relates an individual's coordinates to all variables. However, what can be proven from this relation is that the highest

paid activity (AFV) is always higher than the average wage for activities throughout the study period and regardless of the qualification's category, and vice versa for the lowest paid activity "IEV".

Figure 5: Graph of activities on the factor plane (F1, F2)



Source: Output of applying the PCA method to the data table using the R program

5. 3.3/ Coordinates of supplementary variables: It seems that the first factor can be called " factor of average wage for the study period ". To confirm this, we calculated the average wage for the period within each activity, and then calculated the linear correlation coefficient between this new variable and the first factor. After this procedure, we actually found a very strong correlation, equal to 0.96. This conclusion can also be supported by positive correlations between this factor and the supplementary variables RC9, RM9, RE9, RC19, RM19, RE19, although these are not very strong correlations (Appendix 4).

5. 4/ Interpretation of the second factor

5. 4.1/ Coordinates of active variables: The variables representing the highest paid qualification category "executives" in all years of study are opposed to all variables representing the lowest paid category "enforcement agents", all of which are correlated with this axis, in particular "executives" (Appendix 4). The difference between these two variables in any given year is a measure of dispersion, that is the range. If the first factor is set as "the average wages of the activity for the study period", this opposition indicates that, in the case of equal averages monthly wages for given activities, it is certain that some of these activities are characterized by high salaries for executives and low salaries for enforcement agents (i.e. the salary range is large), while others of these activities, on the contrary, are characterized by low salaries for executives and high salaries for enforcement agents (i.e., the salary range is small). It should be noted here that the comparison is not made between the salaries of the executives and the salaries of the enforcement agents, but each qualification's categories compares with itself between these activities.

5.4.2/ **Coordinates of individuals (activities):** Based on the transition rules, and as we explained in the previous paragraph, we will inevitably find positive coordinates for activities with a large wage range, and negative coordinates for activities with a small wage range. The activity of "Real estate, rental and services to enterprises in the private sector "ISEV" has the largest coordinate (4,47) on this factor, and thus the wages in this activity are characterized by a large disparity between the wages of executives and the enforcement agents, as executives receive high wages and pay the enforcement agents very weak wages, and the same is true for the activity of trade and repair in the private sector "CRV" and the activity of health in the private sector "SV". In return the activity of real estate, rental and services to enterprises in the public sector "ISEP" has the smallest coordinates on This factor (-3.18), but it is not represented by the first factor (the average wage for the study period) because the value of COS^2 related to it is almost zero (0.01), so it cannot be said with certainty that the explanation of this opposition applies to it, and it is better to avoid interpreting it as such. On the contrary, we note that the activities (Table 2): manufactured Industries in the public sector "IMP", financial activities in the public sector "AFP", trade and repair in the public sector "CRP" are represented by the first factor, and have the smallest coordinates on the second factor, and thus these activities are characterized by low wages for executives category if compared with their peers in activities that are similar in average wages for the period of study (ISEV, SV, CRV), and with high wages for enforcement agents, meaning that these activities have the least disparity in wages between executives and enforcement agents among the activities with high averages wages.

5. 4.3/ Coordinates of supplementary variables: The supplementary variables R2C9, R2E9, R2C19, R2E19 representing the proportion (share) of the qualification categories "executives" and "enforcement agents" of the total average salary of all qualifications within a given activity in a given legal sector at the beginning and end of the study period are located on the positive upper side concerning executives (a strong positive correlation with F2), and on its lower negative side are the enforcement agents (a medium negative correlation with F2), which is another indicator that this axis represents a factor of opposition for high-wage activities for executives and low wages for enforcement agents (large wage dispersion) and for relatively high-paying activities for executives and high wages for enforcement agents (wage dispersion is small or reasonable), provided that all such opposing activities are similar in average wages for the study period.

To ensure that the second axis represents the extent of the dispersion of wages for the study period, we calculated the linear correlation coefficient between the average wages for the study period (the average wages for the executives in the study period for a given activity – the average wages of the enforcement agents in the study period for that activity) and the second factor (F2), and we actually obtained a value of almost equal to one (0.993). It can also be observed the opposite position on this axis of the two modalities (Public, Private)of the supplementary qualitative variable that represents the legal sector, which indicates that the activities similar in average wages for the study period are divided into two groups, one characterized by a large extent of dispersion and belongs to the private legal sector, and another characterized by a small and reasonable dispersion and belong to the public sector, that is, the sector also plays a role in the extent to which wages are dispersed in a group of activities.

5. 5/ Interpretation of the first factorial plane (F1,F2)

5. 5.1/ Linear correlations between variables: Projecting the variables point cloud on the first factorial plane Preserve 91.42% of the information contained in the primary data table, which is represented by the total inertia (total variance) of this cloud (or individuals point cloud). On the other hand, the vectors representing these 18 variables and the angles formed by these vectors are only slightly deformed, which can be neglected given their original dimension (18 dimensions). From the graph of the variables (correlation circle), we can see that the ends of

projected vectors representing the variables do not touch the correlation circle but are very close to that. Thus, we can read and determine the intensity of linear correlations between all variables geometrically, without fear of falling into misleading the projected convergences of variables' vectors and the angles formed by these vectors. Reading and determining the intensity of linear correlations is done by observing the measurement of the angles formed by the vectors of the variables, where if the measurement of an angle is approaching zero, this indicates that there is a strong positive correlation between the two variables whose vectors form this angle, because the cosine of an angle whose measure approaches zero is equal to a value approaching one, and in the same way we read and determine the intensity of linear correlations in the case of an angle measuring close to $\pi/2$, the latter indicates that there is a weak correlation, Because $\cos(\pi/2)=0$, and in the case of an angle approaching π, the correlation is strong and negative, because $\cos(\pi)=-1$.

Using this property, linear correlations between active variables can be presented as follows:

- There are strong to very strong positive linear correlations between any qualification category and itself in the different years two by two, because the angle formed by any two vectors of the same qualification in two different years is between zero and 30 degrees (for example: C15, C17, M15, M19, E13 and E19), and this is very logical because the evolution of the wages of the qualification categories is approaching monotony, especially since we found the first factor can be called the average wages of the activity for the period of study, as well as representing a size factor.

- There are strong to very strong positive correlations between qualification categories: control agents and enforcement agents in any year two by two, because the angle formed by any two vectors of these qualifications in any given year is between 25° and 45° (for example: E11, M17, E19, M15).

- There are positive and average correlations between the qualification categories: Executives and control agents in any given year, two by two (except C13 and M11, which are strongly positive correlated), because the angle formed by any two vectors of these qualifications in any year, except C13 and M11, is between 50° and 65° (e.g. C11, M17, C15 and M13).

- There are weak to very weak positive correlations between the qualification categories: Executives and enforcement agents in any given year (e.g. E9, C13, E11 and C17), because the angle formed by any two vectors of these qualification in any given year is between 70° and 90°.

5. 5.2/ Naming of Artificial Variables (Factors): Clearly, after applying the standard principal components analysis method to the primary data table, and the interpretation of the two principal factors (artificial variables) that make up the first factorial plane(F1, F2), we almost completely summarize 18 variables that represent the evolution of wages for qualification categories from 2009 to 2019 by activities, whether in the public or private sectors, into only two artificial variables (two principal components) as follows:

- We name the first factor: The average wage of the activity for the period of study.

- We name the second factor: The extent of the dispersion of wages between executives and enforcement agents within the activity.

5. 5.3/ Classification of activities: According to the names given to the two factorial axes constituting the first factorial plane, the more an activity in a legal sector with a low monthly average net wage, the more it is positioned to the left of the factorial plane, and vice versa, and the more the extent to which its wages are dispersed between executives and enforcement agents, the higher it is positioned on the factorial plane, and vice versa. We observe, for

example, that the two low-paying activities, namely construction in the private sector "CV" and extractive industries in the private sector "IEV" are positioned close to the first factor axis and at the far left, while they are hardly characterized by the second factor axis.

The distribution of activities on this plane from left to right (Fig. 5) shows us the possibility of distinguishing 5 groups (categories) of activities, which can be described as follows:

Group 1: Activities with very low average wages for all qualification categories: includes extractive industries activity in the private sector "IEV" and construction activity in the private sector "CV". This plane represents it well (cos^2 related to either of them and the plane is very close to 1 (Table 2)).

Group 2: Activities with low to medium wages for all qualification categories, and reasonable dispersion between executives and enforcement agents: activities include IMV, TCV, HRV, HRP, CP, EGEP, SCSP, ISEP, all well represented by the first plane except "CV", poorly represented (cos^2=0.22), and "SCSP", "HRV" represented in an average manner (cos^2=0.54 each). But even if the quality of representation is poor for an individual, its coordinates can be explained (Brigitte Escofier J. P., 2008, p. 41). Thus, even construction activity in the public sector "CP" is not well represented at this factor plane, with coordinates on both axes approaching zero, indicating that this activity is characterized by near-average wages.

Group 3: activities with high average wages, a large dispersion between executives and enforcement agents, and belonging to the private sector: includes CRV, ISEV, SV, all of which are well represented by this factor plane.

Group 4: Activities with high average wages, a reasonable dispersion between executives and enforcement agents, and belonging to the public sector: includes IMP, CRP, AFP, all of which are well represented by this factor plane.

Group 5: Activities with very high average wages for all qualification categories: includes financial activities in the private sector "AFV" and transport and communications activity in the public sector "TCP", which are well represented by this factor plane. The supplementary individual "the activity of extractive industries in the public sector (IEP) can be added at the head of this group, even if it does not contribute to the formation of the factorial axes that make up this factor plane, but the latter represents it well on the one hand, and we know that it is characterized by very high wages for all categories of qualification, and surpasses all other activities in public and private sectors on the other hand.

6. Analysis and interpretation of the results of applying the hierarchical ascending clustering method (HAC)

To support the results of the classification of activities as produced by applying the PCA method to the primary data table, or exploring the activities with another clearer and richer classification of the possible groups that can be formed, we applied the hierarchical ascending clustering method on the two artificial variables (coordinates of activities on the factors F1, F2), and this is by virtue of the fact that the first factor plane explains a large percentage of the total inertia on the one hand, and on the other hand, our goal of applying this method for classification is to put it as a supplement to the analysis with principal components and not separately.

6. 1/ The working principle of the method: The general principle of constructing an ascending hierarchical tree is simple, where at first we have **n** individuals (18 activities in our study) called the final elements. We first measure the distance between them tow by tow, and for this we used the Euclidean distance, which is also used in the principal components analysis method, so the joint interpretation of a hierarchical tree and a factorial plane implies that the resemblance

between two individuals is defined in the same way in both methods. Second, the process of assembling the elements begins according to a certain criterion (Min (single) Linkage, Max (complete) Linkage, Average Linkage, Ward Linkage, ...etc), here we have chosen the Ward's criterion.

The "Ward" algorithm is based on the following principle:

In step **K**, we combine two elements ("individual or group of individuals" with "individual or group of individuals"), moving from a partition composed of **n-K + 1** classes to a partition composed of **n-K** classes. Ward's idea is to choose at each step the grouping of classes such that the increase in intra inertia is minimal. In other words, at any stage of the classification (clustering) process, two classes are chosen to be grouped if the increase (height) in intra-group variance is as small as possible. The increase in intra-group variance (resulting from the grouping of classes **i** and **j** is equal to:

$$\boldsymbol{\sigma(i\,,j)} = \frac{\boldsymbol{m_i m_j}}{\boldsymbol{m_i + m_j}}\,\boldsymbol{d^2(g_i, g_j)}$$

This is a criterion minimized at each step and defines the level index of the nodes of the hierarchy.

where: $\boldsymbol{\sigma(i\,,j)}$**:** The value of the increase in intra-group variance (within-class inertia) resulting from the grouping of classes **i** and **j**.

$\boldsymbol{m_i}$: Total weights of Class **i** individuals. $\boldsymbol{m_j}$: Total weights of class **j** individuals.

$\boldsymbol{d^2(g_i, g_j)}$: Euclidean distance squared between $\boldsymbol{g_i}$ (The centre of gravity of the Class i points) and $\boldsymbol{g_j}$ (The centre of gravity of the Class j points).

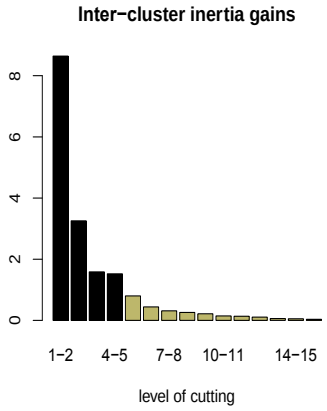
Each time two elements (classes) are grouped, we get a node for them on the hierarchical clustering tree. The height of the node is related to the similarity between the two elements, and is called the level index. This indicator in the "Ward" method represents the value of the increase in intra-group variance, where the sum of the level indexes from the first node to the last node (n-1 nodes) is equal to the total inertia of the cloud, i.e:

The total inertia = $\sum_{\sigma=1}^{n} \sigma_k$

Clusters are visually represented in a hierarchical tree called a dendrogram. The reading of the hierarchical tree is as follows: the less high you have to go in the tree to connect two individuals (activities), and the greater the relationship between these two individuals. (i.e., their monthly wages are close). Thus, the tree highlights, for example, a close relationship between the wage curves for HRP and TCV activities, while the difference is significant between the wage curves for HRP and CRP activities (Figure 7).

6. 2/ The selected partition from hierarchical ascending classification tree: We can cut the tree at a certain level to get a partition. In general, this cutting is done at a level where the groups that make up the partition are interpretable. It also depends on the shape of the tree, as it is better to cut at a level where its branches are long enough, for example, in our study it is possible to cut at a number of groups of 3 or 5. There are also statistical criteria that can be used to determine the level of cut, and we have based one of the most important of these criteria, which is the criterion of the amount of loss in "between-class inertia" (or the amount of gain in "within-class inertia" (Fig. 6)). We can stop grouping classes when the jump is small due to the change in the number of classes from K to K-1, where it means that we are retrieving very little information and therefore there is no point in increasing another grouping of classes.

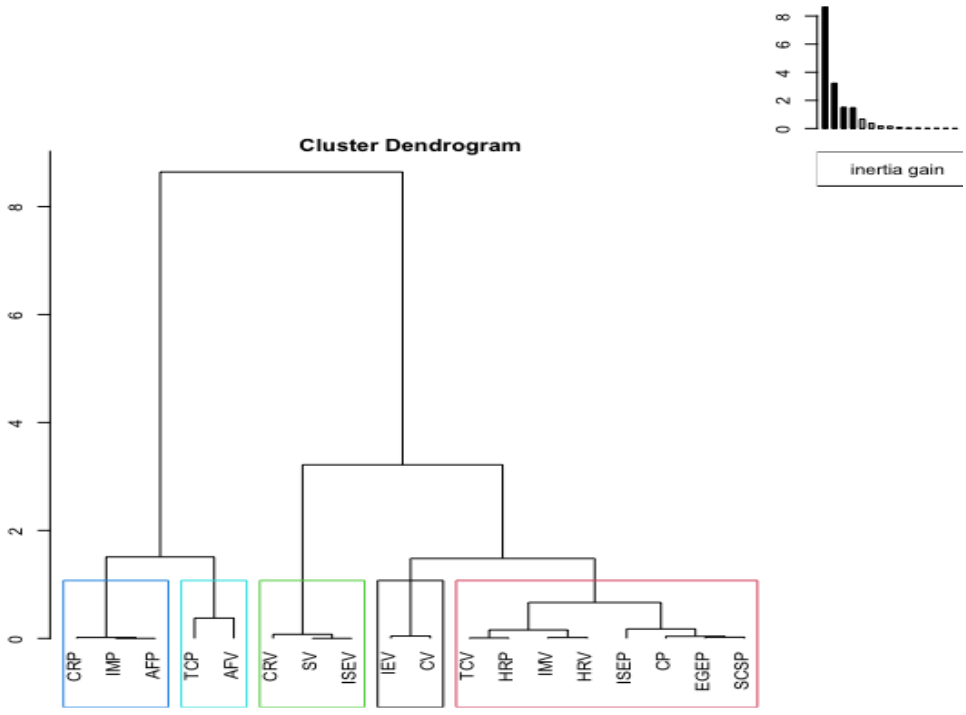
Figure 6: Graph of loss amounts in "between-class inertia"



Source: Output of applying HAC method to the two principal components using the R software

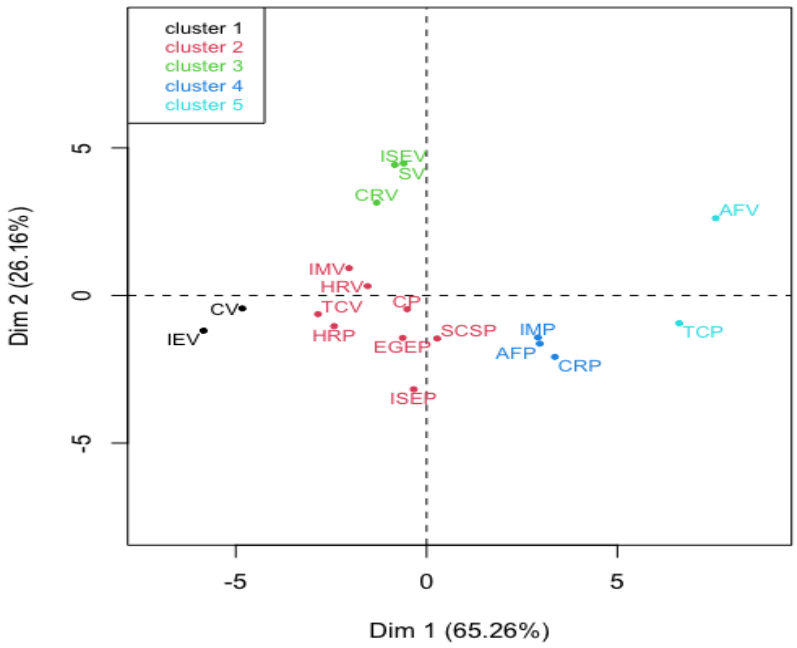
We can see from Figure (6) that the amount of loss of "between-class inertia" if we move from two classes to one is very large, so we prefer to keep two classes or groups. This loss decreases in terms of its amount, but it remains relatively large until the amount of loss related to moving from 5 to 4 groups, immediately afterwards, we notice that the value of this loss as a result of jumping from 6 to 5 sets is small enough, so we stop at the partition consisting of 5 groups, shown in Figure (7).

Figure 7: the hierarchical clustering tree of activities (Dendrogram)



Source: Output of applying HAC method to the two principal components using the R software

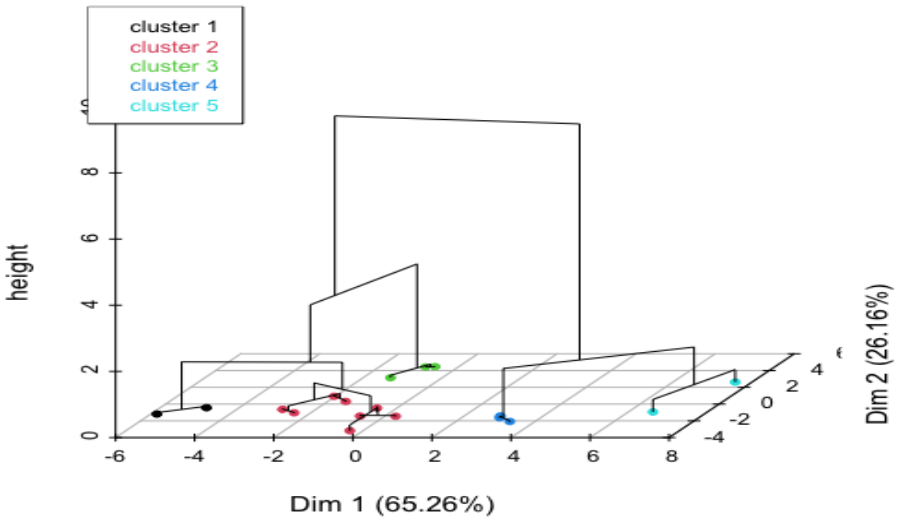
Figure 8: The distribution of the five groups taken in the clustering of activities on the principal factor plane



Source: Output of applying HAC method to the two principal components using the R software

The number of groups determined according to the criterion of the amount of loss in "between-class inertia" in our case fully corresponds to the criterion that interests us the most, which is to take the number of explainable groups, as we did this by analyzing and interpreting these groups using the principal component analysis method.

Figure 9: Three-dimensional graph of the partition taken on the principal factor plane



Source: Output of applying HAC method to the two principal components using the R software

6. 3/ Characterization of groups

The results of applying the hierarchical ascending clustering method to the first and second principal components correspond with the results of the classification of activities as produced by applying the principal component analysis method to the primary data table. It remains only to ensure that these results converge by knowing which factors and variables characterize the partition on the one hand, and which of them characterize the groups that make up this partition. We summarize all this in Table (3), in addition to showing the statistical criteria used to confirm the validity of the description and characterization of the partition and its constituent groups.

6.3.1/ Statistical criteria used to characterize the partition and its groups:

We use three criteria to characterize the partition and its five constituent groups:

- **Correlation ratio "$\hat{\eta}^2$" (Eta 2):** To measure the intensity of correlation between a quantitative variable (in our study, each quantitative variable, whether active or supplementary, as well as either of the two principal components, the first or second) and a qualitative variable (the category or group variable, we have five groups that make up the partition taken for characterization). The closer it is to 1, the stronger we say that the correlation between the two variables is, and the closer it is to 0, the weaker the correlation. $\hat{\eta}^2 = \frac{\sum_{j=1}^{K} n_j (C_j - \bar{X})^2}{\sum_{i=1}^{n} (C_i - \bar{X})^2}$

$\sum_{i=1}^{n} (C_i - \bar{X})^2$: Total Variance. $\sum_{j=1}^{K} n_j \left(C_j - \bar{X}\right)^2$: Variance between groups.

n_j: number of individuals of group j. K: number of groups. n: number of individuals.

C_j: mean of the variable in group j. $\bar{X}$: average of the variable in the sample.

Under the assumption of the independence of the two variables, the statistic $\frac{(n-K)\hat{\eta}^2}{(K-1)(1-\hat{\eta}^2)}$ follows the "Fisher" distribution with degrees of freedom (K-1) and (n-K).

- **Critical probability associated with the critical value of the calculated chi-squared statistic**: to test the independence between two qualitative variables (in our study, between the qualitative variable **SJ** with two modalities "Public" and "private" and the category or group variable with 5 modalities (each set represents a modality).
- **V. test:** to measure the degree of characterization of a group from partition with a variable (the absolute **V. test** greater than 1.96 means that the variable characterizes the group, and the greater the absolute value, the greater the characterization of the variable for the group) on the one hand, and the direction of this characterization on the other hand (a negative value means that the average of a particular quantitative variable within a given group of partition is less than the general average of that variable for all activities, and vice versa, a positive value means that the average of this variable within this group is greater than its overall average for all activities).

$V.test = \frac{(\bar{X}_q - \bar{X})\sqrt{n_q}}{s} \sqrt{\frac{n-n_q}{n-1}}$: Where

$\bar{X}_q$: average of the variable in group q.

$\bar{X}$: the general mean of the variable.

s: Standard deviation of the variable for all activities.

n_q: Number of activities in group q

6. 3.2/ Summary of groups characterization: The characterization and description of the partition and its constituent groups are summarized in the following table:

Table 3: characterization of the partition taken and its constituent activity groups

Group 1	Group 2	Group 3	Group 4	Group 5
1/ Group elements (group activities)				
IEV, CV	IMV, HRV, TCV, HRP, CP, EGEP, SCSP, ISEP	ISEV, SV, CRV	IPM, AFP, CRP	AFV, TCP
2/ Description of the partition by the factor axes: The two axes characterize the partition, where the value of "Eta 2" is 0.95 for the first axis and 0.78 for the second, with a critical probability for each of the two values much less than 0.05 (Appendix 5)				
3/ Description of the partition by the variables: At the level of α = 0.05, all active variables characterize this partition, where E11 is the most characterized with a value of "Eta 2" = 0.95. Supplementary quantitative variables: RC9, RC19, RM9, RM19, RE9, RE19 characterize this partition at a maximum significant level = 0.16. The qualitative variable also characterizes it at a significant level of 0.075 associated with the critical value of the "chi-square" statistic (Appendix 6)				
4/ Description of groups by factorial axes (Appendix 7)				
It is characterized by the first factor, with a V.test = 3,01	It is characterized by both factors, with V.test = 1.65 and -1.45 respectively	It is characterized by the second factor, with a V.test = 3,41	It is characterized by the second factor, with a V.test = -1,47	It is characterized by the first factor, with a V.test = -2.27
5/ Description of groups by active variables (Appendix 8)				
It is characterized by All active variables with positive signs for V.test, of which the strongest is M15 = 3,01 and the weakest is C9= 1,55	It is characterized by All variables that represent the categories "Control agents" and "enforcement agents" with positive signs for V.test, of which the strongest is E9 = 2,3	It is characterized by All variables that represent the category "Executives" with positive signs for V.test, and by all variables that represent the category "enforcement agents" with negative signs.	It is characterized by All variables that represent the category "Executives" with negative signs for V.test, of which the strongest is C13 = -2.14 and the weakest is C11= -1.86	It is characterized by All active variables with negative signs for V.test, of which the strongest is M12 = - 2.57 and the weakest is C15= -1.53
5/ Description of groups by supplementary variables (Appendix 8)				
It is characterized by RM19 and RC19 with positive signs for (V.test), which are equal to	It is characterized by R2C19, with a negative sign for (V.test), which is equal to -2,02, and	It is strongly characterized by all the supplementary variables that	It is characterized by R2M9, with a positive sign for (V.test), which is equal to 1,66, and	Only R2E9 characterizes it moderately, and with a positive sign for (V.test), which

2.35 and 2 respectively	by the qualitative variable "SJ" with its "private" modality.	represent the category "Executives" with positive signs for V.test, and by the qualitative variable "SJ" with its "private" modality.	RE9 with a negative sign, wich is equal to -1,66.	is equal to 1.49.

Source: Personal setup based on the output of applying the hierarchical clustering method to F1,F2 using the R program

Conclusion

In this study, we examined a multidimensional exploration of wage evolution in Algeria for the period 2009-2019 by economic activities in both the legal sectors, public and private. To answer the problematic of how wages are distributed temporally between activities and whether or not different groups can be formed, we chose two complementary methods of multidimensional exploratory statistical analysis, namely principal components analysis and hierarchical clustering.

After analyzing and interpreting the results of both methods, two factors were extracted summarizing 18 variables that make up the primary data table with an information retention rate representing the total variance of the activities point cloud (or variables) estimated at (91.41%). Thus, it is possible to express without redundancy of information the evolution of wages of economic activities in their public and private sectors chronologically (from 2009 to 2019) and in different categories of qualification with only two artificial variables (two principal components or two principal factors), namely: the average wage of the activity for the period of study period, and the extent of the dispersion of wages between executives and enforcement agents within the activity.

The two main hypotheses of the study were also confirmed, the hypothesis " Wages are distributed chronologically between activities at an increasing pace, but in a way that is close to monotony in terms of rates of increase ", we found that it was achieved through the size factor that characterizes the first factor axis, while the hypothesis " More than two clearly different groups of activities can be formed throughout the study period.", we found that it was also achieved through the size factor of the first factor axis, which highlighted three groups of activities that differ mainly among themselves in the total average wages for the period of each activity, and through the second factor, which highlighted two groups of activities, which differ mainly among themselves in the extent to which their wages are dispersed between two categories of qualification, executives and enforcement agents. In addition, this partition of activities has been confirmed through another complementary method to this one, which is the hierarchical ascending clustering method. We apply the latter to the two principal components, the first and second, by virtue of the fact that they maintained a large proportion of the information. The following are the activities that make up the groups and their characteristics:

- The first group: activities with very low average wages for all categories of qualification: includes extractive industries and construction in the private sector.
- The second group: activities with low to medium average wages for all qualification categories, and a reasonable dispersion between executives and enforcement agents, includes: real estate, rental and services to enterprises in the public sector, construction in the public sector, Social and personal community services in the public sector, production and distribution of electricity, gas and water in the public sector, hotels and restaurants in the public sector,

hotels and restaurants in the private sector, transport and communications in the private sector, manufactured industries in the private sector.

- The third group: activities with high average wages, a large dispersion between executives and enforcement agents, and belonging to the private sector: health, real estate, rental and services to enterprises, trade and repair.

- Group IV: Activities with high average wages, a reasonable dispersion between executives and enforcement agents, and belonging to the public sector, includes: financial activities, trade and repair, manufactured industries.

- Group V: Activities with very high average wages for all categories of qualification, includes in order, extractive industries in the public sector, financial activities in the private sector, and transport and communications activity in the public sector.

Appendices

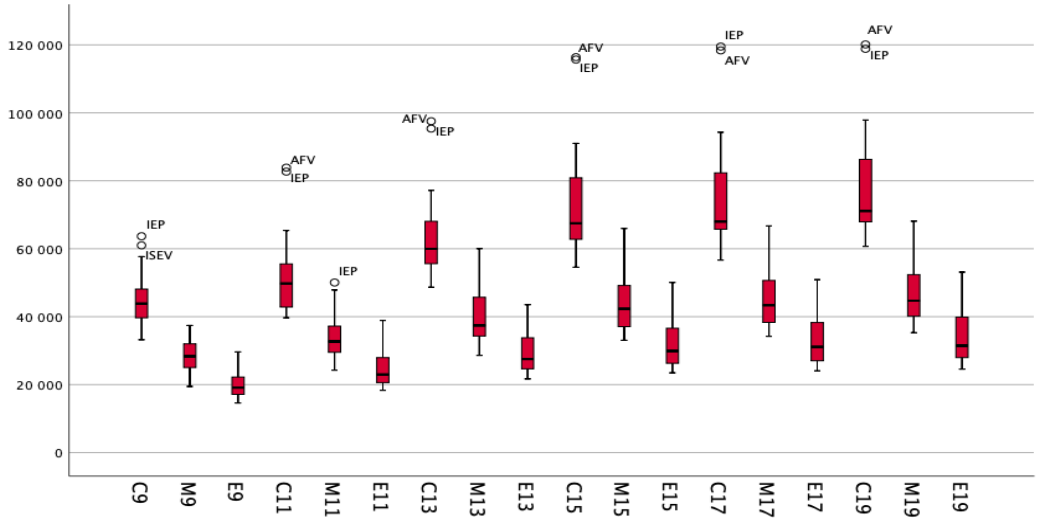
Appendix 1: Definition of Variables and Individuals

Note : QA : " Qantitative variable" , QS : " supplementary variable", QLS : qualitative variable

Individuals (activities)		Variables		
National public sector		**Symbol**	**Variable name**	
symbol	**Individual Name**	**C9**	Average salary of executives in 2009 for a given activity	(QA)
IEP*	Extractive Industries	**M9**	Average salary of control agents in 2009 for a given activity	QA
IMP	Manufactured Industries	**E9**	Average remuneration of enforcement agents in 2009 for a given activity	QA
EGEP	Production and distribution of electricity, gas and water	**C11**	Average salary of executives in 2011 for a given activity	QA
CP	Construction	**M11**	Average salary of control agents in 2011 for a given activity	QA
CRP	Trade and repair	**E11**	Average remuneration of enforcement agents in 2011 for a given activity	QA
HRP	Hotels & Restaurants	**C13**	Average salary of executives in 2013 for a given activity	QA
TCP	Transport & Communications	**M13**	Average salary of control agents in 2013 for a given activity	QA
AFP	Financial Activities	**E13**	Average remuneration of enforcement agents in 2013 for a given activity	QA
ISEP	Real estate, rental and services to enterprises	**C15**	Average salary of executives in 2015 for a given activity	QA
SCSP	Social and personal community services	**M15**	Average salary of control agents in 2015 for a given activity	QA
National Private Sector		**E15**	Average remuneration of enforcement agents in 2015 for a given activity	QA
IEV	Extractive Industries	**C17**	Average salary of executives in 2017 for a given activity	QA
IMV	Manufactured Industries	**M17**	Average salary of control agents in 2017 for a given activity	QA
ISEV	Real estate, rental and services to enterprises	**E17**	Average remuneration of enforcement agents in 2017 for a given activity	QA
CV	Construction	**C19**	Average salary of executives in 2019 for a given activity	QA
CRV	Trade and repair	**M19**	Average salary of control agents in 2019 for a given activity	QA
HRV	Hotels & Restaurants	**E19**	Average remuneration of enforcement agents in 2019 for a given activity	QA
TCV	Transport & Communications	**RC9**	Ratio of wages of a given activity to the total average wages of executives for all activities in 2009	QA
AFV	Financial Activities	**RM9**	Ratio of wages of a given activity to the total average wages of control agents for all activities in 2009	(QS)
SV	Health	**RE9**	Ratio of wages of a given activity to the total average wages of enforcement agents for all activities in 2009	QS
		RC19	Ratio of wages of a given activity to the total average wages of executives for all activities in 2019	QS
		RM19	Ratio of wages of a given activity to the total average wages of control agents for all activities in 2019	QS
		RE19	Ratio of wages of a given activity to the total average wages of enforcement agents for all activities in 2019	QS
		R2C9	ratio of the average salary of the executives category to the overall	QS

		average salary, all categories combined, (in 2009)	
	R2M9	ratio of the average salary of the control agents category to the overall average salary, all categories combined, (in 2009)	QS
	R2E9	ratio of the average salary of the enforcement agents category to the overall average salary, all categories combined, (in 2009)	QS
	R2C19	ratio of the average salary of the executives category to the overall average salary, all categories combined, (in 2019)	QS
	R2M19	ratio of the average salary of the control agents category to the overall average salary, all categories combined, (in 2019)	QS
	R2E19	ratio of the average salary of the enforcement agents category to the overall average salary, all categories combined, (in 2019)	QS
	SJ	The legal sector, has two modalities: public and private	QLS

Appendix 2: Graphical representation of active variables after the first winsorizing



Source: In-person setup based on the initial data table and using SPSS software

Appendix 3: Eigenvalues and the proportions of variance explained by each axis

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6.
Eigenvalues	11.747	4.709	0.665	0.298	0.215	0.153
Variance % of va	65.259	26.161	3.692	1.658	1.193	0.852
Cumulative % of var.	65.259	91.420	95.112	96.770	97.963	98.815

Source: Output of applying the PCA method to the data table using the R program

Appendix 4: Correlations (coordinates) between active, supplementary variables and factors

Coordinates of active variables						Coordinates of supplementary variables		
Variable	Factor1(F1)	Factor 2 (F2)	Variable	(F1)	(F2)	Variable	(F1)	(F2)
C9	0,44	0,77	C15	0,54	0,81	RC9	0,16	0,89
M9	0,9	-0,09	M15	0,97	-0,09	RM9	0,2	0,8
E9	0,76	-0,44	E15	0,88	-0,42	RE9	0,52	0,34
C11	0,61	0,71	C17	0,53	0,8	RC19	0,38	0,85

M11	0,93	0,19	M17	0,96	-0,14	RM19	0,38	0,66
E11	O,93	-0,31	E17	0,88	-0,43	RE19	0,61	0,27
C13	0,69	0,68	C19	0,51	0,82	R2C9	-0,4	0,72
M13	0,98	-0,07	M19	0,96	-0,15	R2M9	-0,34	0,11
E13	0,85	-0,49	E19	0,87	-0,46	R2E9	-0,39	-0,41
						R2C19	-0,43	0,76
						R2M19	-0,39	0,02
						R2E19	-0,35	-0,51

Source: Output of applying the PCA method to the data table using the R program

Appendix 5 : Link between the cluster variable and the Factors

	Eta2	P-value
Dim.1	0.9523534	1.836348e-08
Dim.2	0.7777422	3.441182e-04

Appendix 6: Link between the cluster variable and the quantitative variables

	Eta2	P-value
E11	0.953	1.6e-8
M13	0.901	0.00000203
M15	0.896	0.00000284
E13	0.895	0.000003
C19	0.889	0.00000423
M17	0.886	0.00000489
M19	0.877	0.00000816
C15	0.867	0.0000131
C17	0.863	0.000016
E15	0.862	0.000017
E17	0.862	0.0000171

	Eta2	P-value
E19	0.859	0.0000193
C13	0.857	0.0000211
M11	0.83	0.0000644
M9	0.758	0.000579
C11	0.757	0.000602
RC19	0.693	0.00257
E9	0.68	0.00328
RE19	0.653	0.00539
R2C19	0.639	0.00683
C9	0.576	0.0179
RC9	0.51	0.0416

Appendix 7: Description of each cluster by quantitative variables ($\alpha = 0,05$)

==========

$`1`

v.test Mean in category Overall mean sd in category Overall sd p.value

Dim.1 -2.269727 -5.336427 -1.079383e-16 0.5078129 3.427335 0.02322418

$`2`

NULL

$`3`

v.test. Mean in category Overall mean sd in category Overall sd p.value

Dim.2 3.410968 4.014221 -6.846375e-16 0.6156536 2.170016 0.00064732

$`4`

NULL

$`5`

v.test Mean in category Overall mean sd in category Overall sd p.value

Dim.1 3.019657 7.099613 -1.079383e-16 0.4782194 3.427335 0.0025306

Appendix 8: Description of each cluster by quantitative variables and categories

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
C11	-1.68	-1.87	1.84	-0.0566	2.51
C13	-1.97	-2.15	2.07	0.494	2.32
C15	-1.53	-2.12	2.72	-0.37	2.09
C17	-1.59	-1.99	2.69	-0.484	2.12
C19	-1.59	-1.91	2.74	-0.684	2.18
C9	-1.75	-1.36	2.15	-0.182	1.56
E11	-2.03	-0.744	-1.66	2.18	2.6
E13	-1.73	-0.45	-2.1	2.25	2.27
E15	-1.71	-0.483	-1.88	1.64	2.76
E17	-1.72	-0.465	-1.91	1.72	2.68
E19	-1.77	-0.368	-1.97	1.82	2.54
E9	-1.74	-0.568	-1.48	2.32	1.64
M11	-2.38	-1.11	-0.202	1.33	2.79
M13	-2.13	-1.18	-0.836	1.82	2.84

M15	-1.79	-1.53	-0.739	1.75	3.01
M17	-1.79	-1.43	-0.87	1.8	2.95
M19	-1.82	-1.35	-0.92	1.78	2.93
M9	-2.57	-0.53	-0.678	1.87	2
R2C19	0.354	-0.174	2.79	-2.02	-0.982
R2C9	0.325	-0.374	2.5	-1.52	-0.901
RC19	-1.16	-1.53	2.39	-1.06	2.01
RC9	-1.02	-1.11	2.52	-1.03	1
RE19	-0.684	-1.23	-0.304	-0.272	3.31
RE9	-0.59	-1.66	0.271	0.452	2.36
RM19	-0.396	-1.61	1.23	-0.74	2.36

Source: Output of applying HAC method to the two principal components using the R software

1. Bibliography List :

2. **1.** ONS. (2014). *Revenus salariaux et caractéristiques individuelles.* Statistique sociale. Office nationale des statistiques, Alger, Algeria.
3. **2.** Boutouba Mohamed, T. A. (2022). Wage's gap between economic sectors in Algeria. Revue Les Cahiers du POIDEX, Abdelhamid Ibn Badis University, Mostaganem, *XI*(2), pp. 273, 297.
4. **3.** Wahiba, S. (2008). An econometric study of wage fairness in Algeria. *Magister thesis*. Hasiba Ben Bouali university ,Chlef, Algeria.
5. **4.** Nourine Mouloud, M. M. (2020). Factors affecting the relationship between education and wages in the service sector, an econometric study applied to Algeria Telecom in Mostaganem. Strategy and development review, Abdelhamid Ibn Badis University, Mostaganem, *I*(6), pp. 135, 154.
6. **5.** Pierre Cahuc, Z. (2001). Le marché du travail. De bock Université Bruxelles, Belgique.
7. **6.** INSEE. (2006). Les salaires en France. Institut national de la statistique et des études économiques. INSEE, Paris.
8. **7.** Tchibozo, G. (1998). *2conomie du travail.* DUNOD, Paris, France.
9. **8.** Teiva, S. (2010). Etude des déterminats de relation positive entre la taille des organisations et les salaires au Canada. Mémoire présenté comme exigence partielle de la maitrise en économie. Université du Québec, Montréal, Canada.
10. **9.** ONS. (2010). *Résultas de l'enquete nationale sur les salaires des entreprises.* Office nationale des statistiques. Alger.

11. **10.** Joseph F. Hair JR, W. C. (2019). Multivariate data analysis. CENGAGE, USA.
12. **11.** Planchon, V. (2005). Traitement des valeurs aberrantes: conceots actuels et tendance générale. Revue Biotechnol.Agron.Soc.Enviro, Université de Liège, Belgique *IX*(1), p. 20.
13. **12.** Ludovic Lebart, M. P. (2006). Statistique exploratoire multidimensionnelle, visualisation et inférence en fouilles de données. DUNOD, Paris, France.
14. **13, 14.** Brigitte Escofier, J. P. (2008). Analyses factorielles simples et multiples, objectifs, méthodes et interprétation. DUNOD, Paris, France.
15. **15.** ONS. (2009). (2011). (2013). (2015). (2017). (2019). Résultas de l'enquete nationale sur les salaires des entreprises. Office nationale des statistiques. Alger, Algeria.